Multi-Scale Convolutional Reconstruction Defense Against Adversarial and Real-World Corruptions in Plant Disease Recognition

MD Ilias Bappi, Urusha Shakhakarmi, Jisoo Shin, Kyungbaek Kim
Department of Artificial Intelligence Convergence
Chonnam National University
Gwangju, South Korea
i_bappi@jnu.ac.kr, urushasakha@gmail.com, wkdltn811@gmail.com, kyungbaekkim@jnu.ac.kr

***Abstract*—Deep learning (DL) has emerged as a vital component of modern precision agriculture, enabling fast and reliable plant disease diagnosis directly from leaf images. Despite this progress, current models remain highly susceptible to adversarial perturbations and environmental corruptions, which can distort fine-grained lesion characteristics such as texture, edge sharpness, and color variation. These failures become especially concerning in real-world agricultural pipelines involving drones, field cameras, and IoT devices, where blur, noise, haze, and compression artifacts are common. To address these challenges, we introduce a multi-scale masked autoencoder (MSAE)–based defense framework integrated with a ConvNeXt V2 classifier to enhance robustness against both digital attacks and natural corruptions. The MSAE is designed to reconstruct lesion structures at multiple spatial resolutions, alleviating the over-smoothing and detail loss observed in conventional single-scale denoisers. By combining adversarial examples (Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), and AutoAttack (AA)) with ImageNet-C style corruptions, our method establishes a unified and realistic robustness evaluation pipeline. Experimental evaluation on a benchmark plant leaf disease dataset demonstrates that while the baseline ConvNeXt V2 model achieves 98.98% accuracy on clean images, its performance drops to 65.22% under combined adversarial and corruption settings. With the proposed MSAE-ConvNeXt defense, accuracy is restored to 85.64%, and the framework achieves a mean robust accuracy (mRA) of 74.98% across FGSM, PGD, and AutoAttack. These results confirm that multi-scale reconstruction significantly strengthens model resilience, offering a promising direction toward dependable and field-ready agricultural AI systems.**

***Index Terms*—Adversarial Robustness, Precision Agriculture, Masked Autoencoder, Plant Disease Classification, Multi-Scale Reconstruction.**

I. Introduction

Plant pests and diseases threaten global food security, causing an estimated 20-40% annual crop loss and over USD 220 billion in economic damage [1], [2]. As agriculture increasingly adopts data-driven practices, DL models deployed on drones, smartphones, and IoT devices have become central to automated leaf disease monitoring [3]–[5]. Modern CNN and transformer architectures now achieve over 96% accuracy on curated datasets such as PlantVillage and PlantDoc [3], [6]. However, these systems are typically trained under clean conditions, whereas real farms present noisy, variable, and adversarial environments that can severely degrade DL performance [7]–[9].

Recent studies have shown that even imperceptible perturbations generated by FGSM, PGD, and related attacks can drastically reduce classification accuracy [10]. You et al. report an 87.6% error rate under a GP-MI-FGSM attack for an EfficientNet-based model [11], demonstrating the fragility of current plant disease classifiers. Such perturbations can flip predictions across visually similar disease categories, degrading treatment decisions and propagating losses across the agricultural supply chain [11], [12]. Beyond adversarial attacks, field imagery suffers from blur, noise, haze, and compression artifacts caused by drone motion, sensor limitations, dust, and IoT bandwidth constraints. Robustness research commonly models these degradations using ImageNet-C corruptions [13], [14], yet corruption-aware evaluation remains underexplored in agricultural AI.

To improve robustness, prior works have explored adversarial training, attention-enhanced architectures, and lightweight compression [11], [12]. Reconstruction-based defenses such as CAEs and denoisers have also been applied, but their single-scale decoders often over-smooth critical lesion structures—blurring edges, suppressing fine necrotic spots, and weakening discoloration cues. Our earlier encoder-based CAE defense partially improved robustness under FGSM and PGD but still failed to preserve multi-scale lesion details [15], [16].

These limitations motivate a defense mechanism that (i) jointly addresses digital adversarial attacks and real-world corruptions, and (ii) reconstructs images in a lesion-preserving, multi-scale manner. In this work, we introduce a robust classification pipeline integrating a multi-scale convolutional masked autoencoder (MSAE) with a ConvNeXt V2 backbone. The MSAE operates as a front-end reconstruction module that restores fine-grained lesion textures, edges, and color variations from images distorted by FGSM, PGD, AutoAttack, or ImageNet-C corruptions. The reconstructed output is then classified by ConvNeXt V2.

Our contributions are summarized as follows:

- We present a unified robustness setting for plant disease classification that jointly evaluates digital adversarial at-

tacks (FGSM, PGD, AutoAttack) and ImageNet-C based real-world corruptions.
- We propose a multi-scale convolutional masked autoencoder capable of reconstructing lesion-preserving features that conventional CAEs fail to recover.
- We develop a two-stage robust pipeline (MSAE + ConvNeXt V2) and demonstrate improved prediction stability across diverse perturbations while maintaining high clean-image accuracy.

II. Related Work

Deep learning has shown strong performance in plant disease classification, with recent studies combining CNNs, transformers, and IoT-based imaging systems to support automated field monitoring [17], [18]. These works highlight that modern architectures can achieve high accuracy under clean conditions but seldom address robustness to real-world noise or perturbations. Adversarial robustness in agricultural vision has only recently gained traction. You et al. demonstrate that gradient-based attacks such as GP-MI-FGSM can induce error rates exceeding 80% on plant leaf classifiers despite strong clean performance [11]. More recent explainability-driven methods incorporate adversarial training or knowledge distillation to improve stability [19], yet these approaches primarily focus on specific attack types and do not consider robustness to environmental corruptions common in drone or IoT image capture.

Reconstruction-based defenses have been explored through denoising networks and autoencoders. Chung, Seyeon, et al. propose a denoising–DenseNet pipeline to handle noisy agricultural images [20], while variational autoencoder designs have been used to enhance interpretability rather than adversarial robustness [21]. These methods, however, rely on single-scale decoding, which often oversmooths fine-grained lesion patterns critical for disease identification. Masked autoencoders have recently been adopted for self-supervised feature learning in crop disease recognition [22]–[24], but they are typically used for representation learning on clean images [25]. Their potential as a front-end defense against both adversarial and corruption-based degradation remains unexplored.

In contrast to the above, our previous encoder-based defense study integrated a CAE with ConvNeXt V2 to mitigate FGSM and PGD attacks on plant leaf images [15]. While this approach partially restored accuracy, qualitative analysis revealed that single-scale CAE reconstruction tends to over-smooth lesion boundaries and suppress fine-grained discoloration patterns. The present work advances this line of research by (i) adopting a multi-scale masked autoencoder specifically designed to preserve lesion structures across scales, and (ii) evaluating robustness under a unified threat model that combines gradient-based adversarial attacks with ImageNet-C style corruptions, thereby addressing a gap left by prior adversarial training, denoising, and MAE-based representation learning approaches.

III. Methodology

This section presents the complete design of our robust plant disease classification framework. We begin with an overview of the full pipeline (Fig. 1), followed by a detailed description of (i) digital adversarial attacks, (ii) real-world corruption modeling through ImageNet-C, (iii) the proposed multi-scale convolutional MAE, and finally (iv) the ConvNeXt V2 classification head. Figure 1 illustrates the two-stage architecture comprising a clean classification pipeline and a robust reconstruction pipeline. In the clean setting, images from the PlantVillage dataset are preprocessed and directly fed into ConvNeXt V2 for disease classification. In the robust setting, test images undergo adversarial perturbations (FGSM, PGD, AutoAttack) and ImageNet-C corruptions before being reconstructed by the proposed multi-scale MAE. The reconstructed image is then forwarded to ConvNeXt V2 for final classification.

A. Digital Adversarial Attacks

Adversarial attacks aim to generate minimally perturbed inputs that mislead the classifier. Given an image x and true label y, an adversarial example x^{adv} is defined as:

$$x^{adv} = x + \delta, \quad \text{s.t. } \|\delta\|_\infty \leq \epsilon, \tag{1}$$

where ϵ controls the perturbation strength. Examples of adversarial samples used in our evaluation are shown in Fig. 2.

1) Fast Gradient Sign Method

FGSM [26] generates a perturbation by taking a single gradient step that maximizes the loss:

$$x_{\text{FGSM}} = x + \epsilon \cdot \text{sign}\left(\nabla_x J(\theta, x, y)\right), \tag{2}$$

where J is the classification loss. FGSM perturbs all pixels in the direction of the gradient, often destroying high-frequency lesion textures.

2) Projected Gradient Descent

PGD [27] extends FGSM with iterative refinement:

$$x_{t+1} = \Pi_{\mathcal{B}_\epsilon(x)}\Big(x_t + \alpha \cdot \text{sign}(\nabla_x J(\theta, x_t, y))\Big), \tag{3}$$

where α is the step size and Π projects the result onto the ϵ-ball. PGD is widely considered the "strongest first-order attack" because it gradually removes lesion-specific information.

3) AutoAttack

AutoAttack [28] is a parameter-free ensemble attack composed of: APGD-CE: Auto-PGD on cross-entropy; APGD-DLR: Auto-PGD on DLR loss; FAB Attack: decision boundary–based perturbation; Square Attack: score-based black-box attack. AutoAttack is considered more reliable than single-step attacks since it searches in multiple perturbation subspaces, producing highly diverse perturbation patterns that emulate real-world noise amplification.

B. Real-World Corruptions: ImageNet-C Adaptation

Digital attacks represent worst-case perturbations, but agricultural images captured via drones and IoT devices often degrade due to physical factors. To simulate these conditions, we apply the full set of 15 ImageNet-C corruption types to the

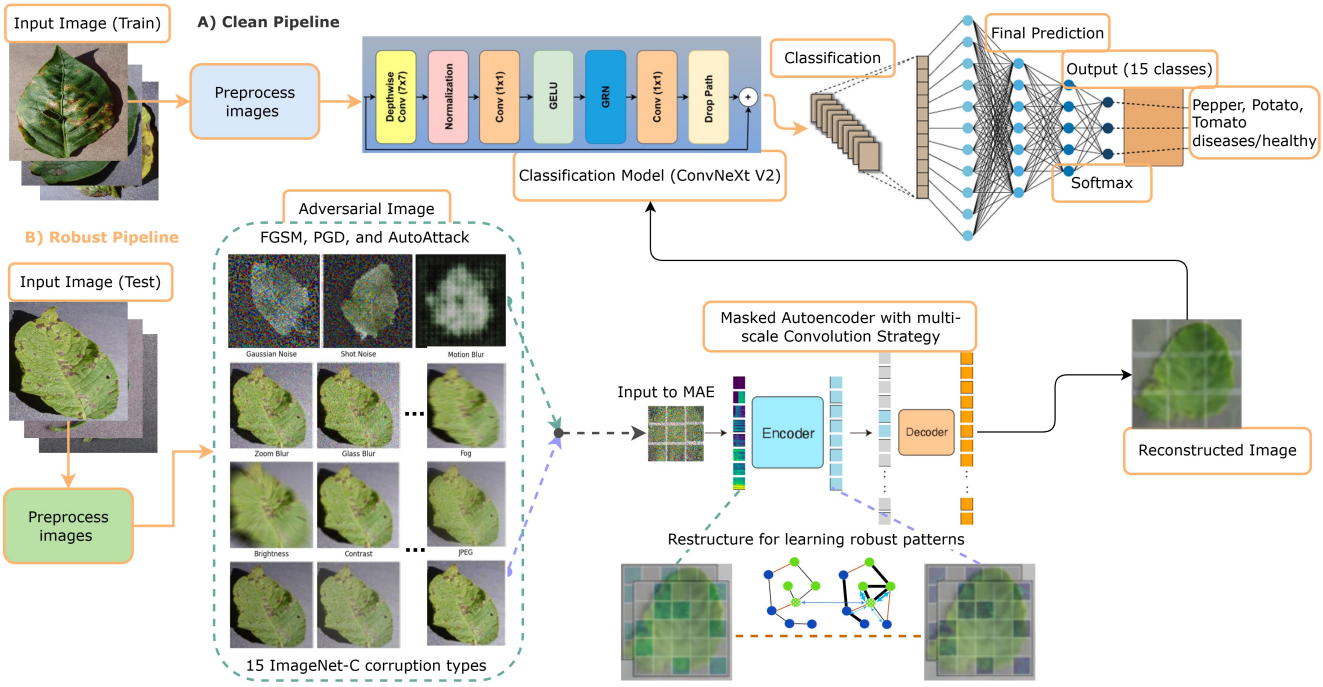

Fig. 1. Overall framework consisting of (A) the clean pipeline using ConvNeXt V2 and (B) the robust pipeline. Test images are perturbed by digital attacks or ImageNet-C corruptions and then reconstructed by a multi-scale masked autoencoder before classification.

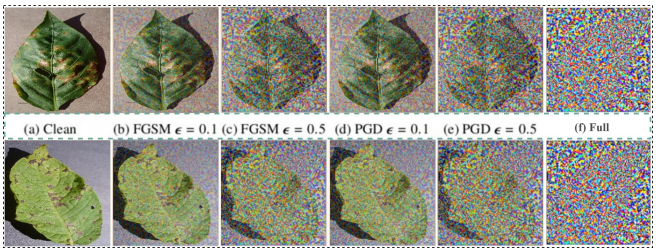

Fig. 2. Examples of adversarial perturbations applied to PlantVillage leaves using FGSM and PGD at multiple ϵ values. Even low-magnitude perturbations distort fine texture, color, and lesion boundaries.

PlantVillage dataset [13]. Given a corruption operator $c_k(\cdot, s)$ of type k and severity s:

$$x_{\text{corr}}^{(k,s)} = c_k(x, s), \quad s \in \{1, \ldots, 5\}. \tag{4}$$

Figure 3 shows examples of noise, blur, weather, and digital corruptions applied to our dataset.

These corruptions degrade lesion boundaries, color contrast, and small-spot structures—making them ideal for evaluating real-world robustness.

C. Multi-Scale Convolutional Masked Autoencoder

The core defense module is a multi-scale convolutional MAE designed to reconstruct lesion-preserving features from corrupted inputs. Its architecture is shown in Fig. 4.

The input image x is divided into patches $\mathcal{P} = \{p_i\}$. A random subset is masked using a binary mask m:

$$\tilde{p}_i = m_i \cdot p_i. \tag{5}$$

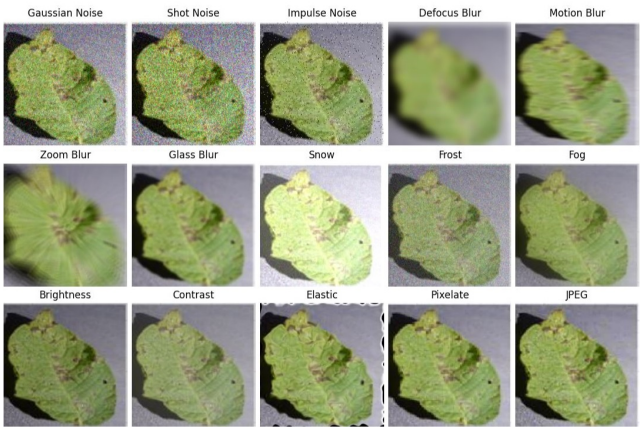

Fig. 3. Examples of ImageNet-C corruptions adapted to PlantVillage leaf images, including noise (Gaussian, shot), blur (defocus, zoom), weather effects (fog, frost), and digital distortions (JPEG, pixelate, brightness).

Masked patches are processed by three convolutional blocks:

$$z_i = E_{\theta_e}(\tilde{p}_i), \tag{6}$$

where each block captures different receptive-field scales:

- Block 1: high-frequency lesion edges
- Block 2: mid-level spot boundaries
- Block 3: large necrotic regions and leaf shape

Transformer blocks then model global dependencies across patches. The Multi-Scale Decoder reconstructs missing patches using:

$$\hat{x} = D_{\theta_d}(z, m), \tag{7}$$

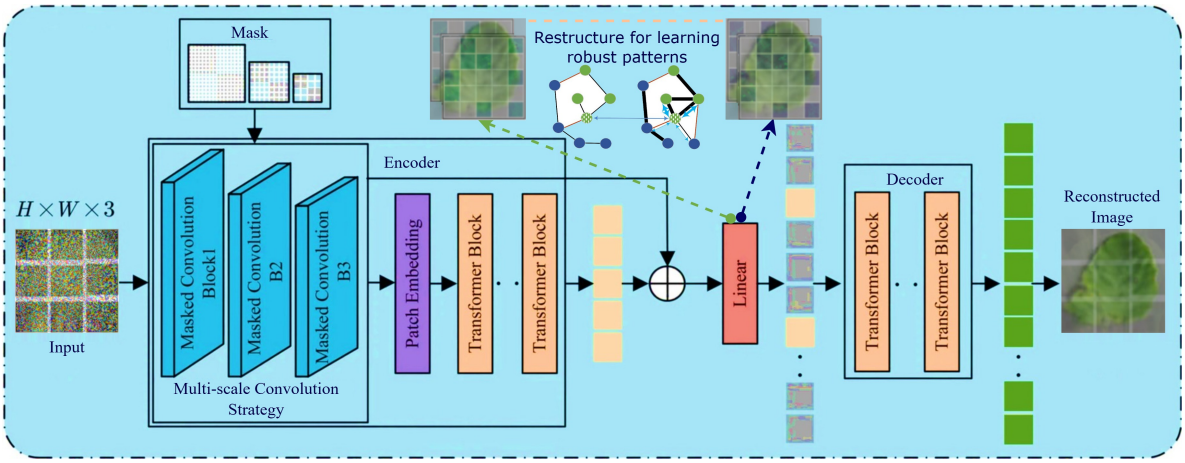

Fig. 4. Architecture of the proposed multi-scale convolutional masked autoencoder. Masked convolution blocks extract robust local features, transformer blocks capture long-range dependencies, and the multi-scale decoder reconstructs high-fidelity leaf structures.

A multi-scale fusion step aggregates features:

$$F_{\text{fuse}} = \sum_{s=1}^{S} W_s F_s, \tag{8}$$

where each F_s corresponds to a specific resolution. The reconstruction loss of MAE is trained with pixel-level L2 loss:

$$\mathcal{L}_{MAE} = \|x - \hat{x}\|_2^2. \tag{9}$$

This enables the model to recover subtle lesions without over-smoothing. Finally, the reconstructed image $\hat{x}$ is forwarded to ConvNeXt V2:

$$h = f_{\text{ConvNeXtV2}}(\hat{x}), \tag{10}$$

followed by a fully connected layer:

$$o = Wh + b, \tag{11}$$

and softmax:

$$\hat{y} = \text{softmax}(o), \tag{12}$$

producing the predicted plant disease class among 15 categories. This separation improves robustness to both digital and real-world corruptions.

IV. Experiment & Analysis

We conducted all experiments on the PlantVillage dataset [29], a widely used benchmark for leaf-based disease classification comprising high-resolution images of pepper, potato, and tomato leaves across 15 classes, including both diseased and healthy samples. Each image was resized to 256×256, normalized to $[0, 1]$, and augmented through random flipping and color jitter during training. For the robustness pipeline, we further generated adversarial samples (FGSM, PGD, AutoAttack) at multiple perturbation magnitudes and applied the full set of 15 ImageNet-C corruptions to simulate real-world acquisition conditions. Model training was implemented in PyTorch using an NVIDIA RTX 4070 GPU with 12 GB VRAM, 32 GB system RAM, and an Intel Core i7-14700KF processor. The clean ConvNeXt V2 classifier was trained for 60 epochs using the AdamW optimizer with a learning rate of 1×10^{-4} and cross-entropy loss. The multi-scale MAE defense module was trained separately for 200 epochs using mean squared error (MSE) reconstruction loss and the same optimizer settings. During inference, the reconstructed image produced by the MSAE was passed through the pre-trained ConvNeXt V2 classifier. To provide a comprehensive evaluation of model robustness, we report standard classification metrics including accuracy, precision, recall, and F1-score for clean, attacked, and reconstructed images. Additionally, following robustness evaluation protocols, we compute the per-attack robust accuracy for FGSM, PGD, and AutoAttack, and report the mean robust accuracy (mRA) to summarize overall resilience against diverse perturbations. We generate perturbations and AutoAttack under an ℓ_∞ threat model. FGSM is applied with $\epsilon \in \{8/255\}$, while PGD uses $\epsilon = 8/255$, a step size of $\alpha = 2/255$. AutoAttack is evaluated using the AA-standard configuration. To model real-world image degradation, we apply all 15 ImageNet-C corruption types at severity levels $s \in \{1, 2, 3, 4, 5\}$. Results are averaged over all corruption types and severities per test image.

A. Results

The baseline ConvNeXt V2 classifier was trained for a maximum of 100 epochs, with the best performance observed at epoch 60, achieving a training loss of 0.0012, a validation loss of 0.0056, and a clean accuracy of 98.98%. This model serves as the reference for evaluating robustness before and after applying our defense module. Figure 5 shows the reconstruction loss (MSE) of the proposed multi-scale MAE over 200 training epochs. The MAE exhibits smooth and consistent convergence, with the training and validation curves stabilizing at 0.0100 and 0.0096, respectively. The validation loss remaining slightly below the training loss indicates strong generalization and effective reconstruction of lesion structures under adversarial and corrupted inputs.

Table I summarizes the performance of the baseline Con-

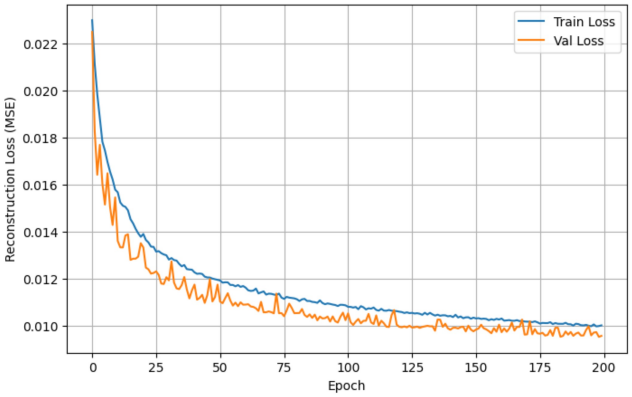

Fig. 5. Reconstruction: both training and validation losses decrease steadily, demonstrating stable convergence and effective reconstruction capability.

vNeXt V2 classifier, the model under adversarial and corruption stress, and the final robust pipeline incorporating the MAE reconstruction module.

TABLE I
COMPARISON OF CLEAN, ATTACKED, AND MAE-ENHANCED ROBUST CLASSIFICATION PERFORMANCE ON THE PLANTVILLAGE DATASET.

Model	**Accuracy**	**Precision**	**Recall**	**F1**
Clean ConvNeXt V2	0.9898	0.9892	0.9895	0.9799
Under Attack/Corruption	0.6522	0.6617	0.6522	0.6470
MSAE + ConvNeXt V2	0.8564	0.8594	0.8464	0.8545

We additionally evaluate robustness using FGSM, PGD, and AutoAttack. The robust pipeline achieves strong per-attack performance and a high mean robust accuracy (mRA): FGSM: 0.8427, PGD: 0.6956, AutoAttack: 0.7112, and mRA: 0.7498. These results validate the effectiveness of the MSAE in restoring perturbed lesion structures prior to classification.

B. Discussion

The clean ConvNeXt V2 model achieves near-perfect performance, confirming the high separability of PlantVillage under ideal conditions. However, applying adversarial attacks and ImageNet-C corruptions causes accuracy to drop by more than 33% (from 0.9898 to 0.6522), highlighting the vulnerability of standard CNN/transformer models to both digital and real-world distortions. Integrating the proposed MSAE significantly mitigates this degradation. The MSAE+ConvNeXt V2 pipeline restores accuracy from 0.6522 to 0.8564, recovering over 20% absolute performance. The high mRA score of 0.7498 further indicates strong robustness across different attack modalities. Overall, the results demonstrate that learning multi-scale lesion-aware reconstructions is highly beneficial for plant disease classification in operational settings such as UAV monitoring and IoT-based crop inspection.

C. Comparison with Recent Adversarially Robust Models

Table II presents a condensed comparison between the proposed MSMAE+ConvNeXt V2 pipeline and recent adversarially aware plant disease classification systems. Here, we report only three essential attributes: model backbone, defense strategy, and the available clean and adversarial robustness metrics.

Compared with prior adversarially-aware models, our MSMAE+ConvNeXt V2 pipeline exhibits substantially stronger robustness across diverse threat settings. SimAM–EfficientNet achieves high clean accuracy but lacks a defense mechanism, resulting in a large accuracy decline under GP–MI–FGSM attacks. Models based on adversarial training and knowledge distillation [19] improve resilience, maintaining approximately 83% accuracy under BIM perturbations. Denoising-based approaches such as RIDNet [**?**] preserve 89–92% accuracy under PGD, while ViT models with FGSM-augmented training sustain nearly clean-level performance under FGSM noise. In contrast, the proposed model is evaluated under a broader and stronger threat model, including FGSM, PGD, AutoAttack, and additionally ImageNet-C corruptions. Despite this wider stress test, our method maintains a mean robust accuracy of 74.98% while retaining a high 98.98% clean accuracy. This demonstrates that multi-scale masked reconstruction effectively preserves fine lesion structures and yields consistent robustness under both digital adversarial attacks and real-world corruptions.

V. CONCLUSION

This work presented a robust plant disease classification framework that combines a multi-scale masked autoencoder with a ConvNeXt V2 backbone to defend against both adversarial attacks and real-world corruptions. While the clean model achieved high accuracy, its performance dropped sharply under FGSM, PGD, and AutoAttack perturbations as well as ImageNet-C distortions. The proposed MSMAE substantially mitigated this degradation by reconstructing lesion-preserving features before classification, restoring more than 20% absolute accuracy and achieving a strong mean robust accuracy across diverse threat settings. In future work, we plan to integrate additional defense strategies, such as adversarial training, frequency-domain reconstruction, and certified robustness techniques. We also aim to incorporate large language models into the pipeline to provide interpretable, human-readable explanations for disease predictions and reconstruction behavior.

ACKNOWLEDGMENT

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry(IPET) through the Agriculture and Food Convergence Technologies Program for Research Manpower development, funded by Ministry of Agriculture, Food and Rural Affairs(MAFRA)(project no. RS-2024-00397026, 34%). This work was supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP)-Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government(MSIT)(IITP-2025-RS-2022-00156287, 33%). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-

TABLE II
COMPARISON OF ADVERSARIAL ROBUSTNESS ACROSS RECENT PLANT MODELS. WHERE: C = CLEAN, R = MRA, F = FGSM, P = PGD, AND A = AA.

Backbone	Defense Strategy	Accuracy / Robustness
EfficientNet + SimAM	No defense; attack (GP–MI–FGSM) [11]	C: 99.31%; robustness not reported; strong drop under attack.
ResNet50 (KD)	Adv. + Knowledge Distillation [19]	C: ≈94.7%; BIM: ≈83.5%.
DenseNet-41	RIDNet denoiser + classifier [30]	PGD ($\epsilon \leq 0.1$): 89.7–92.0%; FGSM: 92.9–98.8%.
ViT-B/16	FGSM-based adv. [31]	FGSM (ϵ 0.01–0.05): >99%; C: 99.4%.
Proposed	**Multi-scale MAE reconstruction**	**C98.98%, R74.98%, F84.27%, P69.56%, A71.12%.**

2025-RS-2023-00256629, 33%) grant funded by the Korea government(MSIT).

REFERENCES

[1] Global Agriculture, "Up to 40 percent of global crop production is lost due to pests and diseases every year: Fao," 2023. Accessed: Nov. 20, 2025.

[2] M. Jung, J. S. Song, A.-Y. Shin, B. Choi, S. Go, S.-Y. Kwon, J. Park, S. G. Park, and Y.-M. Kim, "Construction of deep learning-based disease detection model in plants," *Scientific Reports*, vol. 13, no. 1, p. 7331, 2023.

[3] A. Ahmad, D. Saraswat, and A. E. Gamal, "A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools," *Smart Agricultural Technology*, vol. 3, p. 100083, 2023.

[4] A. Upadhyay, N. S. Chandel, K. P. Singh, S. K. Chakraborty, B. M. Nandede, M. Kumar, A. Subeesh, K. Upendar, A. Salem, and A. Elbeltagi, "Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture," *Artificial Intelligence Review*, vol. 58, no. 3, p. 92, 2025.

[5] J. Yao, S. N. Tran, S. Garg, and S. Sawyer, "Deep learning for plant identification and disease classification from leaf images: Multi-prediction approaches," *ACM Computing Surveys*, vol. 56, no. 6, pp. 1–37, 2024.

[6] A. Y. Ashurov, M. S. A. M. Al-Gaashani, N. A. Samee, R. Alkanhel, G. Atteia, H. A. Abdallah, and M. S. A. Muthanna, "Enhancing plant disease detection through deep learning: A depthwise cnn with squeeze and excitation integration and residual skip connections," *Frontiers in Plant Science*, vol. 15, p. 1505857, 2025.

[7] D. J. Richter, M. I. Bappi, S. S. Kolekar, and K. Kim, "A systematic review of the current state of transfer learning accelerated cnn-based plant leaf disease classification," *IEEE Access*, 2025.

[8] I. Bappi, D. J. Richter, and K. Kim, "Assessing the effectiveness of augmentation techniques in enhancing plant leaf disease classification," *Smart Media Journal*, vol. 14, no. 1, pp. 17–25, 2025.

[9] I.-A. Researchers, "Iiit-a researchers develop ai tech for real-time crop disease detection in indian farms," 2025. Accessed: Nov. 20, 2025.

[10] W. Shafik, A. Tufail, C. D. S. Liyanage, and R. A. A. H. M. Apong, "Using transfer learning-based plant disease classification and detection for sustainable agriculture," *BMC Plant Biology*, vol. 24, no. 1, p. 136, 2024.

[11] H. You, Y. Lu, and H. Tang, "Plant disease classification and adversarial attack using simam-efficientnet and gp-mi-fgsm," *Sustainability*, vol. 15, no. 2, p. 1233, 2023.

[12] P. R. Verma, D. Pantola, and N. P. Singh, "Dynleafnet: A dynamic lightweight architecture for plant disease classification using dynamic residual network with explainable artificial intelligence," *Engineering Applications of Artificial Intelligence*, vol. 155, p. 110937, 2025.

[13] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *arXiv preprint arXiv:1903.12261*, 2019.

[14] T. Saikia, C. Schmid, and T. Brox, "Improving robustness against common corruptions with frequency biased models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10211–10220, 2021.

[15] M. I. Bappi, T. Park, and K. Kim, "An encoder-based defense mechanism against adversarial attacks in plant leaf disease classification," in *Proceedings of the Annual Conference of the Korea Information Processing Society (KIPS)*, pp. 256–259, Korea Information Processing Society, 2025.

[16] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16133–16142, 2023.

[17] A. Jafar, N. Bibi, R. A. Naqvi, A. Sadeghi-Niaraki, and D. Jeong, "Revolutionizing agriculture with artificial intelligence: Plant disease detection methods, applications, and their limitations," *Frontiers in Plant Science*, vol. 15, p. 1356260, 2024.

[18] M. H. Tunio, J. P. Li, X. Zeng, A. Ahmed, S. A. Shah, H. Shaikh, G. A. Mallah, and I. A. Yahya, "Advancing plant disease classification: A robust and generalized approach with transformer-fused convolution and wasserstein domain adaptation," *Computers and Electronics in Agriculture*, vol. 227, p. 109574, 2024.

[19] S.-V. Echim, I.-M. Tăiatu, D.-C. Cercel, and F. Pop, "Explainability-driven leaf disease classification using adversarial training and knowledge distillation," *arXiv preprint arXiv:2401.00334*, 2023.

[20] S. Chung, H. Zhou, D. M. S. Arsa, S. Kim, and H. Kim, "A multi-stage ensemble framework for classifying pig vocalizations under noisy animal farm environments," *Scientific Reports*, vol. 15, no. 1, p. 34703, 2025.

[21] H. Habaragamuwa, Y. Oishi, and K. Tanaka, "Achieving explainability for plant disease classification with disentangled variational autoencoders," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 107982, 2024.

[22] Y. Wang, Y. Yin, Y. Li, T. Qu, Z. Guo, M. Peng, S. Jia, Q. Wang, W. Zhang, and F. Li, "Classification of plant leaf disease recognition based on self-supervised learning," *Agronomy*, vol. 14, no. 3, p. 500, 2024.

[23] M. I. Bappi, D. J. Richter, and K. Kim, "Mae-fvit: Fine-tuned masked autoencoder vision transformer for potato leaf disease classification," in *Proceedings of the 13th International Conference on Smart Media & Applications*, pp. 56–60, 2024.

[24] M. Prasannakumar and K. Latha, "Plant disease identification using contextual mask auto-encoder optimized with dynamic differential annealed optimization algorithm," *Microscopy Research and Technique*, vol. 87, no. 3, pp. 484–494, 2024.

[25] M. I. Bappi, D. J. Richter, H. Jin, and K. Kim, "Sclrnet: A unified self-supervised representation learning framework for comprehensive tuber crop disease recognition," *Journal of Plant Diseases and Protection*, 2025. Revised manuscript submitted (R1), Manuscript ID: JPDP-D-25-00901R1.

[26] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[27] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.

[28] F. Croce and M. Hein, "Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 2206–2216, PMLR, 2020.

[29] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 215232, 2016.

[30] A. Raza, A. H. Pitafi, M. K. Shaikh, and K. Ahmed, "Optimizing potato leaf disease recognition: Insights from densenet-121 and gaussian elimination filter fusion," *Heliyon*, vol. 11, no. 3, 2025.

[31] E. K. Gulsoy, S. Ayas, E. B. Kablan, and M. Ekinci, "Enhancing the adversarial robustness in medical image classification: Exploring adversarial machine learning with vision transformers-based models," *Neural Computing and Applications*, vol. 37, no. 12, pp. 7971–7989, 2025.